DOCUMENT RESUME

ED 189 123 TM 800 288

AUTHOR Mathis, William
TITLE Evaluating: The Policy Implications.
PUB DATE Apr 80
NOTE 14p.: Paper presented at the Annual Meeting of the American Educational Research Association (64th, Boston, MA, April 7-11, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Bias: Educational Policy: Elementary Secondary Education: Evaluation Methods: Information Utilization: *Political Influences: *Program Evaluation: Program Improvement: Self Evaluation (Groups): Values
IDENTIFIERS *Evaluation Problems: External Evaluation

ABSTRACT

Whether initiated by law, regulation, or administrative direction, evaluations have political purposes. Improvement, the classic purpose of evaluation, is most faithfully observed when the importance, funding, and constituency of a program are small. If something is wrong, or if the program's existence is threatened, evaluation can become a weapon in policy disputes, hiding values from constituents in a mystique of scientific inquiry. Purposes may also be reflected in the biases of those who initiate and conduct evaluations, in the selection of a program and objectives of evaluation, and in the amount of evaluation funding relative to program funding. Similarly, problems inherent to evaluation methods are sources of bias. Traditional quantitative measures do not lend themselves to broad and sweeping social programs, such as bilingual education. Finally, evaluation results are often used selectively to further political ends. In conclusion, evaluation can be viewed as an historical enterprise which seeks to recreate the past with selected emphases or biases. (CP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED189123

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

EVALUATING: THE POLICY IMPLICATIONS

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

W. Mathis

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

WILLIAM MATHIS

NEW JERSEY STATE DEPARTMENT OF EDUCATION
TRENTON, NEW JERSEY

TM 800 288

PRESENTED AT AERA ANNUAL MEETING
BOSTON, MASSACHUSETTS
APRIL 7-12, 1980

EVALUATING: THE POLICY IMPLICATIONS*

In Arcadia, evaluations are systematic, rational and orderly endeavors which thoroughly investigate programs, make recommendations for improvement and eliminate programs which have served their purpose or which are ineffective. Public monies are then redirected to more needy endeavors and taxes are reduced.

It does not happen this way.

All educational or social programs are political creations and as such, have a base of vested interests and strong polities. Policy makers have invested some degree of personal interest in proposing and selling the program and are consequently somewhat protective of their commitments. Program personnel have professional stakes in their activities as well as a very real personal concern if they feel their jobs are threatened. Educational programs also serve particular constituencies. Shifts in funding or services would logically be opposed by people who are losing money while being championed by those who wish to gain. The larger and the more powerful the constituency, the less likely a change will occur regardless of the results of any evaluation (19, 2).

The Policy and Political Purposes for Evaluations

Whether evaluations are initiated by law, regulation or administrative direction, or whether the evaluation is periodic or one-time; evaluations are ultimately based in policy and political considerations.

A loose typology of these purposes is here proposed. The five types of functions are divided into program improvement functions and maintenance oriented functions. Any given evaluation will likely be based on a combination of purposes:

---

*The views expressed in this paper are not necessarily the views of the New Jersey State Department of Education.

A. Program improvement functions -

1. Program review and improvement - This function is the classic purpose of evaluation. This idealized form is benign and cooperative in nature but is often used as an overt explanation for evaluations which are conducted for substantially different purposes. Virtually all evaluations claim this purpose as the primary one.

   Generally, this type of purpose is most faithfully adhered to when the program being evaluated (1) is of little consequence, (2) involves a small amount of funds and (3) does not have an extensive constituency. This purpose is also best served in low-profile and incremental change strategies.

   When evaluation consequences are more severe this rational evaluation model will be placed under greater stress.

2A. The *"something's wrong"* purpose - Policy makers and higher echelon staff often receive reports that a particular program is not working as well as it should for some reason or the other. Evaluation activities are then used to find out what is wrong and what should be done about it.

   Such activities may vary in method from a somewhat quiet internal check to an externally based task force.

   The objective is essentially fact-finding and corrective in nature.

3A. Self-improvement behaviors - Social agencies put considerable effort and investment in assuring their own continued existence (1, 19). A program must systematically receive feedback on how well it is meeting user needs and adjust elements in order to assure user satisfaction. If the program is not responsive, then its existence is threatened. This category is characterized by program personnel desires to have a responsive and useful program.

B. Program protection functions - Variants on types two and three are also found in protective evaluations. Any given evaluation may include both improvement and protective purposes.

2B. The *"something's wrong"* purpose - This variant of the something's wrong category is used to protect, maintain or enhance a program other than the one being evaluated.

   Within a limited pool of funds or when there is competition between programs, evaluation can and is used as a weapon in policy disputes (17, 19).

The program objectives which are being evaluated may or may not be the avowed purposes of the program and they may or may not be legitimate aims for the type of program under review (5, 19).

3B. Self-improvement behaviors - This variant of the program improvement purpose is characterized by protective behaviors where *"an evaluation"* is used to conceal program faults or shortcomings.

If a program is threatened, then demonstration that internal or external evaluations are being conducted and that improvements are being made will tend to strengthen the program area against criticisms.

Normally, carefully selected *"external"* evaluators will be called in by program related personnel to check on the problems. The approach may serve as an effective strategen to the degree that the evaluators are credible (1).

4. Legitimizing activities - If a program area is under high threat, then an external expert is often called upon to evaluate the proposed or existing policies in order to assure the internal and external critics that the approach being taken or proposed is a valid course to follow.

Evaluations of staff and reorganizations rely on this evaluative purpose by using external people who are safely living on the opposite side of the continent. This approach is frequently used prior to the implementation of a controversial project which is threatening to program personnel.

5. Constituency building - Whenever policy makers or vested interest groups either wish to promote an idea they think has merit or is in the best interests of that group, an evaluation of current activities is conducted. Thus the evaluation becomes a means to an end (13). Evaluation reports in these cases often have conclusions in the following form:

"Based on these evaluation findings, it was found that we need:

-80,000 more vocational positions (or)
-Community schools in every urban area (or)
-A high school for the artistically talented (or)
-1,500 more guards to protect the teachers

The greatest danger of such approaches is that they can cloak value positions in the mystique of scientific inquiry (16). Such arguments, however, can be quite persuasive in building public support, and constituencies.

A common approach in this policy use of evaluation is the *"pilot project."* Very few project evaluations report that the program was poorly developed, badly implemented and was a sorry idea in the first place. Generally, the project evaluations show that the program is quite successful and must be spread throughout the rest of the district/state/nation. Virtually all hypotheses are supported as the evaluation design assures that they will be confirmed. If the project was not completely successful, this is most frequently attributed to *"lack of enough time," "insuffisient resources"* or similar types of explanations (6, 19).

Sources of Policy or Political Bias Prior to Evaluation Activities

The purpose of initiating evaluations should be but are not necessarily explicit. The types of purposes noted in the previous section may also be reflected in the form of biases in the establishment of the evaluation activity. Bias, in this context, refers to boundaries placed upon an evaluation activity by the manner in which the evaluation was structured. Such biases are inherent tendencies which may or may not be fulfilled. While these biases may reflect a particular purpose, they may or may not be an explicit or conscious purpose. Nevertheless, administrators, evaluators and audiences should be aware of these biases and take them into proper consideration when conducting or reviewing evaluation findings.

1. What group generated the evaluation?

Many programs have evaluation requirements built into their authorizing law or regulation. In such cases, program improvement is generally (but not always) the aim. If the evaluation was generated as a result of criticism from external constituencies, the evaluation is subject to bias based on organizational self-protection or as a weapon in a policy direction dispute. The intensity of the bias and the tendency of the evaluation will vary as a function of the relative power of the program as contrasted with the power of the critics.

Internally ordered evaluations against external criticism will tend toward protectiveness while internal agency criticisms will tend towards internal policy disputes. These tendencies do not rule out legitimate *"something's wrong"* perspectives.

2. Why was this particular program selected for evaluation?

With limited resources and a broad range of programs which could be evaluated, policy makers and higher echelon staff selected the particular area. This selection itself indicates a bias.

If the evaluation is required under law or regulation, the provision itself reflects a bias as contrasted with other areas where evaluation is not mandated.

The evaluation clause may represent questions about the program by opponents and support was either obtained or neutralized by the inclusion of an evaluation clause. *"Sunset"* clauses also represent a polity concerned about extensive government. Again, the positive virtue of periodic review cannot automatically be ruled out.

If the evaluation was directed by the administration, then external criticisms or *"something's wrong"* variants will tend to be the motivating forces.

3. What are the objectives being evaluated? What is included or excluded?

Any broad based program will have a number of different objectives. Those announced by the program personnel are often broader than what can reasonably be expected to be accomplished. The broader objectives were necessary in order to gain the necessary support for the establishment of the program. Thus it is highly likely that any given program will have some failed objectives (2, 4).

The natural evolution of programs will also change the objectives from those espoused in the original aims of the program. First, the translation from goals to program objectives and operational principles will not and cannot be accomplished with absolute fidelity. Secondly, new conditions and demands that have arisen will modify program goals and objectives. The selection of objectives to be evaluated will obviously influence the outcome of the evaluation.

4. Who conducts the evaluation?

A primary consideration is whether the evaluation is conducted internally or externally.

Internal evaluations will tend to be biased toward the protection of the agency. Program failures will more likely be ascribed to lack of external support, limited funds, and insufficient time. Internal organizational pressures may mute criticism of the administrative agency.

More objective evaluations are more likely to be provided by external agencies. However, bias may easily be manifest in the evaluating organization's vested interest and/or the perspectives and biases of the evaluator.

5. How well supported is the evaluation activity?

The match of evaluation resources to the magnitude of the evaluation task also interjects bias into the evaluation. Examination of a $100 million compensatory program with a $7,000 budget cannot produce much more than a cursory review.

Large expenditures on evaluation of a small project can also interject bias in a number of ways such as overwhelming the program, detailed examination of all objectives, and the expenditure of large evaluation resources on one project to the neglect of others. Such disproportionate attention may also reflect on the purposes of the evaluation.

Thus, before a project is directed to evaluators, certain approaches and avenues may have been precluded and the operational space will have been defined. These purposes and bias may have an inordinate effect upon the findings of the evaluation.

Limitations and Biases Within the Evaluation Itself

While this paper is not centrally concerned with problems inherent to evaluation processes, such problems affect the communication of evaluation findings to policy-makers and the utility of the results for policy purposes. A primary consideration may be codified in an axiom:

The precision and accuracy of the evaluation will vary directly with the insignificance and narrow breadth of the program being evaluated.

Broad and sweeping social programs with a great number of interactive and complex components simply do not lend themselves to precise measurement or clean designs. Evaluation in such circumstances cannot be absolutely conclusive due to evaluation methodology limitations. For example, bilingual programs are still in confusion over the basic purpose of whether the program is transitional or maintenance oriented. While the State may have a clear statement on the matter, this message may not have reached the classroom

level. Measurement instruments may not be readily available for the various languages. A traditional testing approach may also be inappropriate because the requirement for services may outlaw the use of control groups. Estimated without treatment expectations interject error in the assessment process and group norms are sadly lacking. If these problems could be overcome, then the question remains as to how much progress is progress. More than likely, the instruments used to measure program accomplishments will not match program objectives.

To compound an already complex example, if sufficient progress was not made, the next problem is determining why. Traditional quantitative measures simply do not have the sophistication to sort out causality. Were the teachers qualified? Were materials available? Student attitudes? Community support? Program administration? Financial support?

Traditional quantitative evaluation methods have been over-sold and often present an empirical mystique which can very often be misleading and give value judgments the sanctity of scientific truths (11, 16, 17).

The limitations of quanitative based evaluations where the method may frequently be elevated above the problem, has led to dissatisfaction with traditional evaluation procedures and a call has been made for qualitative evaluations. While the quality of the program is generally the most critical question, judgments about quality raise new concerns and interject new biases (14, 3, 7, 15).

The primary problem is deciding what is good quality or a desirable end.

Evaluators, working from a particular administrative or statistical base, for example, will select certain portions of a program to examine and will present selected portions of the data according to their training

and biases. There is a natural tendency for statisticians to rely on assessment data while administrators will select management data. More subtle biases will enter based upon how the evaluator thinks a program should run. Colleague oriented persons will view educational processes differently than will autocratically oriented persons.

Administrators may also apply criteria such as orderliness and lack of disturbance while teachers may value activity. Two entirely different evaluation results based on the same activity could occur.

Program expectations will also vary according to each level of responsibility in the administrative hierarchy.

Such judgmental approaches increase the likelihood of unreliable data and place a premium on clear definition and communication of expectations to all involved parties and extensive training of program evaluators to assure an accurate evaluation. Contamination of observations by particularly vocal or biased persons or atypical conditions at the time of evaluation must also be guarded.

Utilization of Evaluation Information in the Policy Process

Evaluation information generally plays only one part and a minor part in program and policy decisions. While rational approaches are valued in the abstract, they do not have a compelling voice as contrasted with vested interests and powerful lobbying agencies (12, 20). The recent study of influences on federal policy by Florio et al. shows that research and evaluation data is ranked low in the decision making process (8).

The use of research data is often selective and is used to further political ends. In New Jersey, a preliminary count of violence and vandalism reports which were received turned up as the lead sentence in an article in the National Education Association's monthly publication. Though the

union representatives were aware that this count included duplicate reports from the phase-in period with a number of false and duplicated filings, the unedited numbers provided greater political force than the smaller corrected number or the emotionally neutral index per thousand students. The assessment report which indicates that 75 percent of the students are mastering basic skills will appear as *"one-fourth flunk minimum levels"* in the press or in the presentations of data for an opposing purpose.

Evaluators tend to be somewhat cautious in filing reports but extensive caveats, footnotes and limitations are ignored as the evaluation results become politicized and brought into play in the political process (5).

As evaluators tend to approach projects in a somewhat narrow and limited sense, they run afoul of policy-makers whose primary concerns are inputs, outputs and *"does it work."* Evaluation results are seldom as clear-cut and unequivocal as policy makers would desire (18, 5, 3).

Conclusions

The history of the word evaluation is the back-formation of an activity. This requires a restructuring against certain values. This paper illustrates a number of ways in which evaluations may be biased. Values of policy makers may have changed between program inception and evaluation, the balance of political forces and constituencies may have shifted (such as the decrease in the proportion of the public who have children in the schools), the program may have shifted purposes in practice, and the program may have simply failed to achieve the desired objectives.

Evaluation can be viewed as an historical enterprise which seeks to recreate the past with certain selected emphases or biases. All

people recreate previous events in their own views and for their own purposes (4). Southern histories of the Civil War vary greatly from Northern histories which vary yet again from a European history of the same conflict. Evaluation is subject to the same process.

"Evaluation is a rational enterprise that takes place in a political context," notes Carol Weiss (19). It is incumbent upon administrators, evaluators and audiences to examine and clearly explicate the purposes, the biases and limitations, the values and shifts in values, and the procedural constraints if the process is to become more rational and if it is to provide more objective data for the improvement of educational programs.

REFERENCES

1. Alkin, Marvin C. and Richard H. Daillak. "Evaluation Utilization." Educational Evaluation and Policy Analysis, Vol. 1, No. 4, July-August, 1979, pp. 41-50.

2. Boyd, William L. "Changing Politics of Curriculum Policy-making for American Schools." Review of Educational Research, Vol. 48, No. 4, Fall, 1978, pp. 577-628.

3. Brickell, Henry. Paper presented in symposium on the Role of State Education Agencies in Educational Research. Annual meeting of the American Educational Research Association, 1979.

4. Button, H. Warren. "Creating More Usable Pasts." History in the Study of Education." Educational Researcher. Vol. 8, No. 5, May, 1979, pp. 3-9.

5. California Educational Management and Evaluation Committee. "The Uses and Limits of Educational Evaluation at the State Level." Unpublished report for California State Department of Education, December, 1978.

6. Cohen, S. Allan and Joan S. Hyman. "How Come So Many Hypotheses in Educational Research are Supported? (A Modest Proposal)." Educational Researcher. Vol. 8, No. 11, December, 1979, pp. 12-16.

7. Eisner, Elliot W. "The Use of Qualitative Forms of Evaluation for Improving Educational Practices." Educational Evaluation and Policy Analysis. Vol. 1, No. 6, Nov.-Dec., 1979, pp. 11-20.

8. Florio, David H., Michael M. Behrman and Diane L. Goltz. "What do Policy Makers Think of Educational Research and Evaluation? Or Do They?" Educational Evaluation and Policy Analysis. Vol. 1, No. 6, Nov.-Dec., 1979, pp. 61-87.

9. Glasman Naftaly S. "A Perspective on Evaluation as an Administrative Function in Education." Educational Evaluation and Policy Analysis. Vol. 1, No. 5, Sept.-Oct., 1979, pp. 39-44.

10. Halperin, Samuel. "The Future and Educational Governance: Prospects and Possibilities." Unpublished paper presented at Summer Institute of Council of Chief State School Officers. July, 1979.

11. House, Ernest R. "Coherence and Credibility: The Aesthetics of Evaluation." Educational Evaluation and Policy Analysis. Vol. 1, No. 5, Sept.-Oct., 1979, pp. 5-18.

12. Iannacone, Lawrence. Paper presented in symposium in Values Imposed by Political Science: Implications for Educational Research and Development. American Educational Research Association Annual Meeting, 1979.

13. Krathwohl, David, R. "The Myth of Value-Free Evaluation." Educational Evaluation and Policy Analysis. Vol. 2, No. 1, Jan.-Feb., 1980, pp. 37-46.

14. Rist, Ray, C. "Blitzkreig Ethnography: On the Transformation of a Method into a Movement." Educational Researcher. Vol. 9, No. 2, Feb., 1980. pp. 8-10.

15. Schubert, William, H. "Recalibrating Educational Research: Toward a Focus on Practice." Educational Researchers. Vol. 9, No. 1, January, 1980, pp. 17-23.

16. Schwab, Joseph. Paper presented in symposium on Values Imposed by the Behavioral and Social Disciplines. American Educational Research Association Meeting, 1979.

17. Stenzel, Norman. "Problems in the Implementation and Acceptance of New Evaluation Approaches in State Departments of Education." Review copy of paper to be presented at American Educational Research Association Meeting, 1980.

18. Tyler, Ralph. Paper presented in symposium on the Role of State Education Agencies in Educational Research. Annual meeting of the American Educational Research Association, 1979.

19. Weiss, Carol, H. "Evaluation Research in the Political Context." Paper presented at the annual meeting of the American Psychological Association, 1973.

20. Wirt, Frederick. Paper presented in symposium on Values Imposed by Political Science. Annual meeting of the American Educational Research Association, 1979.